

# Shark Tank

*Shark Tank* is a reality TV show. Contestants present their idea for a company to a panel of investors (a.k.a. "sharks"), who then decide whether or not to invest in that company. The investors give a certain amount of money in exchange for a percentage stake in the company ("equity"). If you are not familiar with the show, you may want to watch part of an episode [here \(http://abc.go.com/shows/shark-tank\)](http://abc.go.com/shows/shark-tank) to get a sense of how it works. You can also search for a clip on YouTube.

The data that you will examine in this lab contains data about all contestants from the first 6 seasons of the show, including:

- the name and industry of the proposed company
- whether or not it was funded (i.e., the "Deal" column)
- which sharks chose to invest in the venture (N.B. There are 7 regular sharks, not including "Guest". Each shark has a column in the data set, labeled by their last name.)
- if funded, the amount of money the sharks put in and the percentage equity they got in return

To earn full credit on this lab, you should:

- use built-in pandas methods (like `.sum()` and `.max()`) instead of writing a for loop over a `DataFrame` or `Series`
- use the split-apply-combine pattern wherever possible

Of course, if you can't think of a vectorized solution, a `for` loop is still better than no solution at all!

```
In [19]: import pandas as pd
import numpy as np
%matplotlib inline
import matplotlib.pyplot as plt
```

## Question 0. Getting and Cleaning the Data

The data is stored in the CSV file `sharktank.csv`. Read in the data into a Pandas `DataFrame`.

```
In [20]: # YOUR CODE HERE
data = pd.read_csv("sharktank.csv")
data.head()
```

Out[20]:

	Season	No. in series	Company	Deal	Industry	Entrepreneur Gender	Amount	Equity	Corcoran	Cuban
0	1.0	1.0	Ava the Elephant	Yes	Healthcare	Female	\$50,000	55%	1.0	NaN
1	1.0	1.0	Mr. Tod's Pie Factory	Yes	Food and Beverage	Male	\$460,000	50%	1.0	NaN
2	1.0	1.0	Wispots	No	Business Services	Male	NaN	NaN	NaN	NaN
3	1.0	1.0	College Foxes Packing Boxes	No	Lifestyle / Home	Male	NaN	NaN	NaN	NaN
4	1.0	1.0	Ionic Ear	No	Uncertain / Other	Male	NaN	NaN	NaN	NaN

There is one column for each of the sharks. A 1 indicates that they chose to invest in that company, while a missing value indicates that they did not choose to invest in that company. Notice that these missing values show up as NaNs when we read in the data. Fill in these missing values with zeros. Other columns may also contain NaNs; be careful not to fill those columns with zeros, or you may end up with strange results down the line.

```

In [21]: # YOUR CODE HERE
# importing pandas package
#import pandas as pd

# filling a null values using fillna()

#data["Corcoran"].fillna("0", inplace = True)
#data["Cuban"].fillna("0", inplace = True)
#data["Greiner"].fillna("0", inplace = True)
#data["Herjavec"].fillna("0", inplace = True)
#data["John"].fillna("0", inplace = True)
#data["O'Leary"].fillna("0", inplace = True)
#data["Guest"].fillna("0", inplace = True)
#data["Harrington"].fillna("0", inplace = True)

data = data.fillna(0)

data

```

Out[21]:

	Season	No. in series	Company	Deal	Industry	Entrepreneur Gender	Amount	Equity	Corcora
0	1.0	1.0	Ava the Elephant	Yes	Healthcare	Female	\$50,000	55%	1.
1	1.0	1.0	Mr. Tod's Pie Factory	Yes	Food and Beverage	Male	\$460,000	50%	1.
2	1.0	1.0	Wispots	No	Business Services	Male	0	0	0.
3	1.0	1.0	College Foxes Packing Boxes	No	Lifestyle / Home	Male	0	0	0.
4	1.0	1.0	Ionic Ear	No	Uncertain / Other	Male	0	0	0.
...	...	...	...	...	...	...	...	...	.
490	6.0	28.0	You Kick Ass	Yes	Children / Education	Female	\$100,000	10%	0.
491	6.0	29.0	Shark Wheel	Yes	Fitness / Sports	Male	\$225,000	8%	0.
492	6.0	29.0	Gato Cafe	No	Uncertain / Other	Female	0	0	0.
493	6.0	29.0	Sway Motorsports	Yes	Green/CleanTech	Male	\$300,000	20%	0.
494	6.0	29.0	Spikeball	Yes	Fitness / Sports	Male	\$500,000	20%	0.

495 rows × 17 columns

Notice that Amount and Equity are currently being treated as categorical variables ( `dtype: object` ). Can you figure out why this is? Clean up these columns and cast them to numeric types (i.e., a `dtype` of `int` or `float` ) because we'll need to perform mathematical operations on these columns.

```
In [22]: # YOUR CODE HERE
data['Amount'] = data.Amount.replace('[\$,]', '', regex=True).astype(float)
data['Equity'] = data.Equity.replace('[\%,]', '', regex=True).astype(float) / 100

print (data['Amount'])
print (data['Equity'])
data
```

```
0      50000.0
1     460000.0
2         0.0
3         0.0
4         0.0
```

...

```
490    100000.0
491    225000.0
492         0.0
493    300000.0
494    500000.0
```

Name: Amount, Length: 495, dtype: float64

```
0      0.55
1      0.50
2      0.00
3      0.00
4      0.00
```

...

```
490     0.10
491     0.08
492     0.00
493     0.20
494     0.20
```

Name: Equity, Length: 495, dtype: float64

Out[22]:

	Season	No. in series	Company	Deal	Industry	Entrepreneur Gender	Amount	Equity	Corcora
0	1.0	1.0	Ava the Elephant	Yes	Healthcare	Female	50000.0	0.55	1.
1	1.0	1.0	Mr. Tod's Pie Factory	Yes	Food and Beverage	Male	460000.0	0.50	1.
2	1.0	1.0	Wispots	No	Business Services	Male	0.0	0.00	0.
3	1.0	1.0	College Foxes Packing Boxes	No	Lifestyle / Home	Male	0.0	0.00	0.
4	1.0	1.0	Ionic Ear	No	Uncertain / Other	Male	0.0	0.00	0.
...	...	...	...	...	...	...	...	...	.
490	6.0	28.0	You Kick Ass	Yes	Children / Education	Female	100000.0	0.10	0.
491	6.0	29.0	Shark Wheel	Yes	Fitness / Sports	Male	225000.0	0.08	0.
492	6.0	29.0	Gato Cafe	No	Uncertain / Other	Female	0.0	0.00	0.
493	6.0	29.0	Sway Motorsports	Yes	Green/CleanTech	Male	300000.0	0.20	0.
494	6.0	29.0	Spikeball	Yes	Fitness / Sports	Male	500000.0	0.20	0.

495 rows × 17 columns

## Question 1. Which Company was Worth the Most?

The valuation of a company is how much it is worth. If someone invests \$10,000 for a 40\% equity stake in the company, then this means the company must be valued at \$25,000, since 40% of \$25,000 is \ \$10,000.

Calculate the valuation of each company that was funded. Which company was most valuable? Is it the same as the company that received the largest total investment from the sharks?

```
In [23]: # YOUR CODE HERE
print (data.loc[data['Amount'].idxmax()])
# Declare a list that is to be converted into a column

# Using 'Address' as the column name
# and equating it to the list
data['Valuation'] = data['Amount']/data['Equity']

data['Valuation'].replace(np.inf, np.nan, inplace = True)

# Observe the result
print (data.loc[data['Valuation'].idxmax()])
#print (data.max())
#data
```



Season	
6	
No. in series	2
7	
Company	AirCa
r	
Deal	Ye
s	
Industry	Green/CleanTec
h	
Entrepreneur Gender	Mal
e	
Amount	5e+0
6	
Equity	0.
5	
Corcoran	
0	
Cuban	
0	
Greiner	
0	
Herjavec	
1	
John	
0	
O'Leary	
0	
Harrington	
0	
Guest	
0	
Details / Notes	Contingent on getting deal to bring to contin
e...	
Name: 483, dtype: object	
Season	
6	
No. in series	1
1	
Company	Zip
z	
Deal	Ye
s	
Industry	Food and Beverag
e	
Entrepreneur Gender	Mal
e	
Amount	2.5e+0
6	
Equity	0.
1	
Corcoran	
0	
Cuban	
0	
Greiner	
0	

```

Herjavec
0
John
0
O'Leary
1
Harrington
0
Guest
0
Details / Notes          with an option for another $2.5 Million for a
n...
Valuation                2.5e+0
7
Name: 421, dtype: object

```

## YOUR EXPLANATION HERE

Compnay Zipz was the most valuable, the copmany that received the largest total investment was Aircar.

## Question 2. Which Shark Invested the Most?

Calculate the total amount of money that each shark invested over the 6 seasons. Which shark invested the most total money over the 6 seasons?

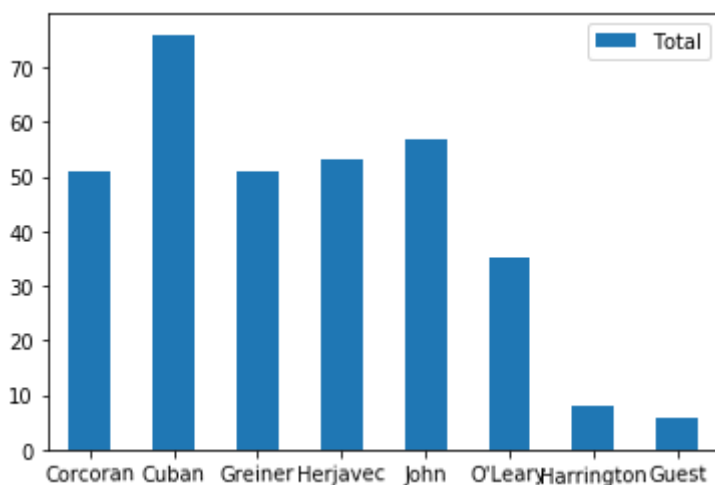
*Hint:* If  $n$  sharks funded a given venture, then the amount that each shark invested is the total amount divided by  $n$ .

```
In [24]: # ENTER CODE HERE.
Total=[data['Corcoran'].sum(),data['Cuban'].sum(),data['Greiner'].sum(),
        data['Herjavec'].sum(),data['John'].sum(),data['O'Leary'].sum(),
        data['Harrington'].sum(),data['Guest'].sum()]

index=['Corcoran','Cuban','Greiner','Herjavec','John','O'Leary','Harrington',
       'Guest',]

print (Total, index)
data2 = pd.DataFrame({'Total':Total},index=index)
#data.sum(axis = 0, skipna = True)
ax=data2.plot.bar(rot=1)
```

```
[51.0, 76.0, 51.0, 53.0, 57.0, 35.0, 8.0, 6.0] ['Corcoran', 'Cuban', 'Greiner', 'Herjavec', 'John', 'O'Leary', 'Harrington', 'Guest']
```



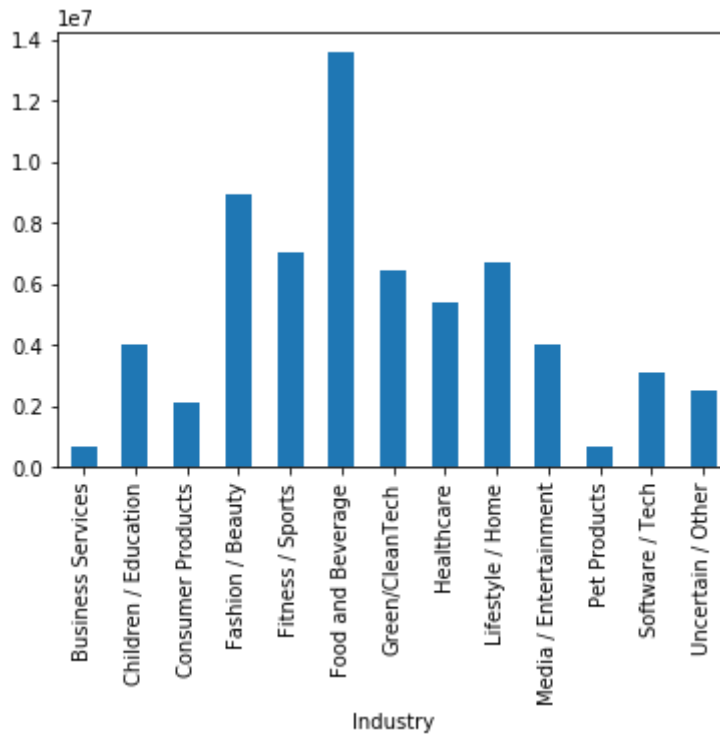
**YOUR EXPLANATION HERE** These results show that Cuban was the shark that invested the most total money.

### Question 3. Do the Sharks Prefer Certain Industries?

Calculate the funding rate (the proportion of companies that were funded) for each industry. Make a visualization showing this information.

```
In [25]: # ENTER CODE HERE.  
#Mean gives us Proportions:  
#(titanic_df.sex == "male").mean()  
#(data.Deal == 'True').mean()  
#(titanic_df.sex == "male") & (titanic_df.survived == 1)  
#data[(data.Deal == "male")].age.plot.hist()  
#data.groupby("Industry").mean()["Deal"].plot(kind='bar')  
df3=data.groupby([ "Company", "Industry"])[ "Amount" ].sum()  
df3  
df4=data.groupby([ "Industry"])[ "Amount" ].sum()  
df4.plot.bar()
```

Out[25]: <matplotlib.axes.\_subplots.AxesSubplot at 0x7f90911c75c0>



**YOUR EXPLANATION HERE**

## Submission Instructions

Once you are finished, follow these steps:

1. Restart the kernel and re-run this notebook from beginning to end by going to `Kernel > Restart Kernel and Run All Cells`.
2. If this process stops halfway through, that means there was an error. Correct the error and repeat Step 1 until the notebook runs from beginning to end.
3. Double check that there is a number next to each code cell and that these numbers are in order.

Then, submit your lab as follows:

1. Go to `File > Export Notebook As > PDF`.
2. Double check that the entire notebook, from beginning to end, is in this PDF file. (If the notebook is cut off, try first exporting the notebook to HTML and printing to PDF.)
3. Upload the PDF and Notebook (ipynb) to iLearn